

```

1 TCGCGGGGAGC CAGAGGGGCC TGCGGTCTCT GGTGGTCTTG CCAGCCCCCTC
51 CTCATCCAGC GGGCTCCGCT GCCTGTGAGG ACTCCCTCAG GTCGGCCACG
101 GGACCTGACG CAACAGGATG GACGAGTCCC CTGAGCCTCT GCAGCAGGGC
151 AGAGGGCCGG TGCCGGTCCG ACGGCAGCGC CCAGCACCCC GGGGTCTGCG
201 TGAGATGCTG AAGGCCAGGC TGTGGTGCAG CTGCTCGTGC AGTGTGCTGT
251 GCGTCCGGGC GCTGGTGCAG GACCTGCTCC CCGCCACGCG CTGGCTGCGT
301 CAGTACCGCC CGCGGGAGTA CCTGGCAGGC GACGTCTATG CTGGGCTGGT
351 CATCGGCATC ATCCTGGTGC CGCAGGCCAT CGCCTACTCA TTGCTGGCCG
401 GGCTGCAGCC CATCTACAGC CTCTATACGT CCTTCTTCGC CAACCTCATC
451 TACTTCTCA TGGGCACCTC ACGGCATGTC TCCGTGGGCA TCTTCAGCCT
501 GCTTTGCCCTC ATGGTGGGGC AGGTGGTGGG CCGGGAGCTC CAGCTGGCCG
551 GCTTTGACCC CTCCAGGAC GGCCTGCAGC CCGGAGCCAA CAGCAGCACC
601 CTCAACGGCT CGGCTGCCAT GCTGGACTGC GGGCGTGA CTACGCCAT
651 CCGTGTGCGC ACCGCCCTCA CGCTGATGAC CGGGCTTTAC CAGGTCCTCA
701 TGGGCGTCC TCGGCTGGGC TTCGTGTCCG CCTACCTCTC ACAGCCACTG
751 CTCGATGGTG TTGCCATGGG GGCCTCCGTG ACCATCCTGA CCTCGCAGCT
801 CAAACACCTG CTGGGCGTGC GGATCCCGCG GCACCAGGGG CCGGCATGG
851 TGGTCTCTAC ATGGCTGAGC CTGCTGCGCG GCGCCGGGCA GGCCAACGTG
901 TCGGACGTGG TCACCAGCAC GGTGTGCCTG GCGGTGCTGC TAGCCGCGAA
951 GGAGCTCTCA GACCGTACC GACACGCTT GAGGGTGCCG CTGCCCACGG
1001 AGCTGTGGTG CATCGTGGTG GCCACACTCG TGTCGCACTT CCGGCAGCTC
1051 CACAACGGCT TTGGCTCGAG CGTGGCTGGC GACATCCCCA CGGGTTTCAT
1101 GCCCCTCAG GTCCAGAGC CCAGCTGAT GCAGCGTGTG GCTTTGGATG
1151 CCGTGGCCCT GGCCCTCGTG GCTGCCGCT TCTCCATCTC GCTGGCGGAG
1201 ATGTTGCCCC GCAGTACGG CTACTCTGTG CGTGCCAACC AGGAGCTGCT
1251 GGCTGTGGGC TGCTGCAACG TGCTACCGCG CTCTCTCCAC TGCTTCGCCA
1301 CCAGCGCCGC CCTGGCCAAG AGCCTGGTGA AGACAGCCAC TGGCTGCCGG
1351 ACACAGCTGT CCAGCGTGGT CAGCGCCACC GTGGTGCTGC TGGTGCTGCT
1401 GCGCTGGCA CCGCTGTTC ACGACCTACA GCGAAGCGTG CTGGCCTGCG
1451 TCATCGTGGT CAGCTGCGG GGGGCCCTGC GCAAGGTGTG GGACCTCCCG
1501 CCGCTGTGGC GGATGAGCCC GGCTGACGCG CTGGTCTGGG CAGGCACCGC
1551 GGCCACCTGT ATGCTGGTCA GCACAGAGGC CCGGCTGCTG GCTGGCGTCA
1601 TCCTCTCGCT GCTCAGCTG GCCGGCCGCA CCAACGCCC ACGCACCGCC
1651 CTGCTGGCCC GCATCGGGGA CACGGCCTTC TACGAGGATG CCACAGAGTT
1701 CGAGGGCCTC GTCCCTGAGC CCGGCGTGCG GGTGTTCCGC TTTGGGGGGC
1751 CGCTGTACTA TGCCAACAAG GACTTCTTCC TGCAGTCACT CTACAGCCTC
1801 ACGGGGCTGG ACGCAGGCTG CATGGCTGCC AGGAGGAAGG AGGGGGGCTC
1851 AGAGACGGGG GTCGGTGAGG GAGGCCCTGC CCAGGGCGAG GACCTGGGCC
1901 CGGTTAGCAC CAGGGCTGCG CTGGTGCCCG CAGCGGCCGG CTTCACACA
1951 GTGGTCATCG ACTGCGCCCC GCTGCTGTTC CTAGACGCAG CCGGTGTGAG
2001 CACGCTGCAG GACCTGCGCC GAGACTACGG GGCCTTGGG ATCAGCCTGC
2051 TGCTAGCTTG GTGCAGCCCG CCTGTGAGAG ACATTCTGAG CAGAGGAGGC
2101 TTCCTCGGGG AGGGCCCGG GGACACGGCT GAGGAGGAGC AGCTGTTCTC
2151 CAGTGTGCAC GATGCCGTGC AGACAGCACG AGCCCGCCAC AGGGAGCTGG
2201 AGGCCACCGA TGTCCATCTG TAGCAGGGCC AGGCTGCCC AGCAGCCTCT
2251 GCTCCCTCCT GGGGACCCAC AGCAGACGTC TGCAAGCCAC TGCTGAGACC
2301 CTTCCAGGG AGGAGCCACC CAAGAGCTGC ACTCTGTGC CACAGCTGCC
2351 CTGGGGAAC CGGGGAACCC CAACTGGGAA AGGAGGCCCT CTGATCACAC
2401 GCAGGACCCA AACATCAGA AATCAAGAAC CTCTGCCTCC GAGCAGGCT
2451 GGCCCAAGT GCTGGCTGGG CCCC AATGCA CCGTCCCTCA GCTCAGAAGG
2501 GATGGGCCCTG ACCTGACGCT CAGGGTTGAC ATCTTATTTG AACAGGGTC
2551 CCCC GCCATC ATGCAGCCTC CAAGGTGCCA AGAGGACTCC CTATGCCAG
2601 GCCTGCCCGG TGCCACCCCT GCTGGTAGGA GCCAGCGGCT CTGGCCAAGT
2651 GCACGAGGGT CTCTGTGTTT CCAGAAGGCC CCACACACCC AAGTGCCCTC
2701 CACACCTCGT GCCTCCCCCT CACAGGTGG CCACCTGCAC CAGCGTCAGG
2751 GCCCAGGGT CTGTGACCGA TGAGACCTCA GCTCAGCCCT CAGGTGCAGT
2801 GGCCCTACCC AGCCTGGCCA GCAGACACAC ACAGGGATGC TCACGGGTGC
2851 ACCAGGAGCC AGGTGCGGCG CAGCCAACCC TGAGCCTGCA GGGAGACCTG
2901 CAGGAAGCCC ACCGTGCCCC ATGCAGGGGC TCCCTCCAGC ACACAGCCCT
2951 CACCCAGCA CAGCCAGCAA GGACACGCTC TCCCAACAG GGTGCTTCGG
3001 CGGGAGGTGG GGAACAAGG GGTCTCCGA GCAGCCCCCA GCCCTCCCTC
3051 CCCATCTGTG CCTCTGTAAG GGGCTCTGGG ACGCCAGAC CCTGCCCGCC
3101 GCCACCTGG TGGTGACAA GCTCCAGCAG CAGTGGGTCC GGACCTGCTT

```

FIGURE 1, page 1 of 3

3151 GATGCCCGCGG TGAGGGACGG CGCCACATA GGCGAGGTTG AGCTGCTGGT  
3201 CCCAGCTGAG GACGTACTGG TCAGCCTGGC TGTGTGGCAG CGGGGGGCTG  
3251 GGGACAACAA AGGGGCGGCT CAGTCCCGAG CCTCAGCATG GCTGGCAGCG  
3301 CGGCTGACAC ACACGTTCAA GCCCAGGACT GCCCGGGCGC AGGATCCAGG  
3351 CGCTGCCCCGT GCGTTCAGTG ACTAATAAAA TGACCCTTAG GGCCAGGAAA  
3401 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAA (SEQ ID NO:1)

**FEATURES:**

5'UTR: 1-117  
Start Codon: 118  
Stop Codon: 2221  
3'UTR: 2224

117 118 2221 2224

# **HOMOLOGOUS PROTEINS:**

## Top BLAST Hits:

	Score	E
CRA 335001098671800 /altid=gi 11545741 /def=ref NP_071325.1  so...	1385	0.0
CRA 335001098639224 /altid=gi 11560117 /def=ref NP_071623.1  su...	1055	0.0
CRA 1000746201930 /altid=gi 6746349 /def=emb CAB69640.1  (AJ223...	654	0.0
CRA 18000004923413 /altid=gi 4557539 /def=ref NP_000103.1  sulf...	649	0.0
CRA 18000004971635 /altid=gi 627422 /def=pir  A54808 diastroph...	649	0.0
CRA 154000124061898 /altid=gi 12054717 /def=emb CAC20729.1  (Y1...	647	0.0
CRA 18000005144885 /altid=gi 6015035 /def=sp O70531 DTD_RAT SUL...	631	e-179
CRA 18000004938377 /altid=gi 6681233 /def=ref NP_031911.1  dias...	622	e-177
CRA 108000024647870 /altid=gi 12730580 /def=ref XP_011158.1  so...	522	e-147
CRA 1000682322799 /altid=gi 6755022 /def=ref NP_035997.1  pendr...	416	e-115

## BLAST dbEST Hits:

	Score	E
gi 10209038 /dataset=dbest /taxon=96...	1015	0.0
gi 7140527 /dataset=dbest /taxon=9606...	769	0.0
gi 5847932 /dataset=dbest /taxon=9606 ...	488	e-135

# **EXPRESSION INFORMATION FOR MODULATORY USE:**

library source (from BLAST dbEST hits):

gi|10209038 Lung  
gi|7140527 Lymph  
gi|5847932 Kidney

## Tissue Screening Panels:

Human heart  
Human Leukocyte  
Thyroid  
Pituitary  
Brain  
Fetal brain  
Adrenal gland  
Testis  
Kidney  
Small intestine  
Pancreas  
Liver  
Lung  
Placenta  
Skeletal muscle  
Spleen  
Hela cells

```

1 MDESPEPLQQ GRGPVPVRRQ RPAPRGLREM LKARLWCSCS CSVLCVRALV
51 QDLLPATRWL RQYRPREYLA GDVMSGVLIG IILVPQAIAY SLLAGLQPIY
101 SLYTSFFANL IYFLMGTSRH VSVGIFSLLC LMVGQVVDRE LQLAGFDPSQ
151 DGLQPGANSS TLNGSAAML D CGRDCYAIRV ATALTMTGL YQVLMGVLRL
201 GFVSAYLSQP LLDGFAMGAS VTILTSQLKH LLGVRI PRHQ GPGMVVLTWL
251 SLLRGAGQAN VCDVVTSTVC LAVLLAAKEL SDRYRHRLRV PLPTELLVIV
301 VATLVSHFGQ LHKRFGSSVA GDIPTGFMPQ QVPEPRMQR VALDAVALAL
351 VAAAFSISLA EMFARSHGYS VRANQELLAV GCCNVLP AFL HCFATS AALA
401 KSLVKTATGC RTQLSSVVSA TVVLLVLLAL APLFHD LQRS VLACVIVVSL
451 RGALRKVWDL PRLWRMSPAD ALVWAGTAAT CMLVSTEAGL LAGVILSLLS
501 LAGRTQRPR T ALLARIGDTA FYEDATEFEG LVPEPGVRVF RFGGPLYAN
551 KDFFLQSLYS LTGLDAGCMA ARRKEGGSET GVGE GGP AQG EDLGPVSTRA
601 ALVPAAAGFH TVVIDCAPLL FLDAAGVSTL QDLRRDY GAL GISLLLACCS
651 PPVRDILSRG GFLGEGPGDT AEEEQ LFLSV HDAVQTARAR HRELEATDVH
701 L (SEQ ID NO:2)

```

#### FEATURES:

##### Functional domains and key regions:

[1] PDOC00001 PS00001 ASN\_GLYCOSYLATION  
N-glycosylation site

Number of matches: 2

1	158-161 NSST
2	163-166 NGSA

[2] PDOC00005 PS00005 PKC\_PHOSPHO\_SITE  
Protein kinase C phosphorylation site

Number of matches: 7

1	117-119 TSR
2	281-283 SDR
3	370-372 SVR
4	449-451 SLR
5	505-507 TQR
6	597-599 STR
7	686-688 TAR

[3] PDOC00006 PS00006 CK2\_PHOSPHO\_SITE  
Casein kinase II phosphorylation site

Number of matches: 7

1	358-361 SLAE
2	467-470 SPAD
3	526-529 TEFE
4	562-565 TGLD
5	629-632 TLQD
6	670-673 TAEE
7	679-682 SVHD

[4] PDOC00007 PS00007 TYR\_PHOSPHO\_SITE  
Tyrosine kinase phosphorylation site

515-522 RIGDTAFY

[5] PDOC00008 PS00008 MYRISTYL  
N-myristoylation site

Number of matches: 15

1	76-81 GLVIGI
2	152-157 GLQPGA
3	156-161 GANSST
4	218-223 GASVTI
5	255-260 GAGQAN

6 316-321 GSSVAG  
7 476-481 GTAATC  
8 489-494 GLLAGV  
9 493-498 GVILSL  
10 563-568 GLDAGC  
11 567-572 GCMAAR  
12 576-581 GGSETG  
13 577-582 GSETGV  
14 581-586 GVGEVG  
15 660-665 GGFLGE

[6] PDOC00012 PS00012 PHOSPHOPANTETHEINE  
Phosphopantetheine attachment site

411-426 RTQLSSVVSATVVLLV

[7] PDOC00870 PS01130 SULFATE\_TRANSP  
Sulfate transporters signature

98-119 PIYSLYTSFFANLIYFLMGTSR

**Membrane spanning structure and domains:**

Helix	Begin	End	Score	Certainty
1	73	93	1.663	Certain
2	98	118	1.558	Certain
3	121	141	0.813	Putative
4	180	200	1.400	Certain
5	209	229	1.017	Certain
6	259	279	1.008	Certain
7	291	311	1.227	Certain
8	344	364	1.585	Certain
9	377	397	1.343	Certain
10	414	434	2.107	Certain
11	483	503	1.446	Certain
12	602	622	0.977	Putative
13	635	655	0.897	Putative

**BLAST Alignment to Top Hit:**

```
>CRA|335001098671800 /altid=gi|11545741 /def=ref|NP_071325.1| solute
  carrier family 26 (sulfate transporter), member 1 [Homo
  sapiens] /org=Homo sapiens /taxon=9606 /dataset=nraa
  /length=701
  Length = 701
```

Score = 1385 bits (3545), Expect = 0.0  
Identities = 698/701 (99%), Positives = 698/701 (99%)  
Frame = +1

```
Query: 1      MDESPEPLQQGRGVPVRRQRPAPRGLREMLKARLWCSCSCSVLCVRALVQDLLPATRWL 180
             MDESPEPLQQGRGVPVRRQRPAPRGLREMLKARLWCSCSCSVLCVRALVQDLLPATRWL
Sbjct: 1      MDESPEPLQQGRGVPVRRQRPAPRGLREMLKARLWCSCSCSVLCVRALVQDLLPATRWL 60

Query: 181    RQYRPREYLAGDVMSGLVIGIILVPQAIAYSLLAGLQPIYSLYTSFFANLIYFLMGTSRH 360
             RQYRPREYLAGDVMSGLVIGIILVPQAIAYSLLAGLQPIYSLYTSFFANLIYFLMGTSRH
Sbjct: 61    RQYRPREYLAGDVMSGLVIGIILVPQAIAYSLLAGLQPIYSLYTSFFANLIYFLMGTSRH 120

Query: 361    VSVGIFSLLCMLVGQVVDRELQLAGFDPSQDGLQPGANSSTLNGSAAMLDCGRDCYAIRV 540
             VSVGIFSLLCMLVGQVVDRELQLAGFDPSQDGLQPGANSSTLNGSAAMLDCGRDCYAIRV
Sbjct: 121    VSVGIFSLLCMLVGQVVDRELQLAGFDPSQDGLQPGANSSTLNGSAAMLDCGRDCYAIRV 180

Query: 541    ATALTIMTGLYQVLMGVLRLGFVSAYLSQPLLDGFAMGASVTILTSQLKHL LGVRIPRHQ 720
             ATALTIMTGLYQVLMGVLRLGFVSAYLSQPLLDGFAMGASVTILTSQLKHL LGVRIPRHQ
Sbjct: 181    ATALTIMTGLYQVLMGVLRLGFVSAYLSQPLLDGFAMGASVTILTSQLKHL LGVRIPRHQ 240

Query: 721    GPGMVVLTWLSLLRGAGQANVCDVVTSTVCLAVLLAAKELSDRYRHRLRVPLPTELLVIV 900
             GPGMVVLTWLSLLRGAGQANVCDVVTSTVCLAVLLAAKELSDRYRHRLRVPLPTELLVIV
Sbjct: 241    GPGMVVLTWLSLLRGAGQANVCDVVTSTVCLAVLLAAKELSDRYRHRLRVPLPTELLVIV 300

Query: 901    VATLVSHFGQLHKRFSGSSVAGDIP TGFMPQVPEPRLMQRVALDAVALALVAAAFSISLA 1080
             VATLVSHFGQLHKRFSGSSVAGDIP TGFMPQVPEPRLMQRVALDAVALALVAAAFSISLA
Sbjct: 301    VATLVSHFGQLHKRFSGSSVAGDIP TGFMPQVPEPRLMQRVALDAVALALVAAAFSISLA 360

Query: 1081   EMFARSHGYSVRANQELLAVGCCNVLP AFLHCFATSAA LAKSLVKTATGCRTQLSSVVSA 1260
             EMFARSHGYSVRANQELLAVGCCNVLP AFLHCFATSAA LAKSLVKTATGCRTQLSSVVSA
Sbjct: 361    EMFARSHGYSVRANQELLAVGCCNVLP AFLHCFATSAA LAKSLVKTATGCRTQLSSVVSA 420

Query: 1261   TVVLLVLLALAPLFHDLQRSVLACVIVVSLRGALRKVWDL PRLWRMSPADALVWAGTAAT 1440
             TVVLLVLLALAPLFHDLQRSVLACVIVVSLRGALRKVW  PRLWRMSPADALVWAGTAAT
Sbjct: 421    TVVLLVLLALAPLFHDLQRSVLACVIVVSLRGALRKVWGFPRLWRMSPADALVWAGTAAT 480

Query: 1441   CMLVSTEAGLLAGVILSLLSLAGRTQRPR TALLARIGDTAFYEDATEFEGLVPEPGVRVF 1620
             CMLVSTEAGLLAGVILSLLSLAGRTQRPR TALLARIGDTAFYEDATEFEGLVPEPGVRVF
Sbjct: 481    CMLVSTEAGLLAGVILSLLSLAGRTQRPR TALLARIGDTAFYEDATEFEGLVPEPGVRVF 540

Query: 1621   RFGGPLY YANKDFFLQSLYSLTGLDAGCMAARRKEGGSETGVGEGGPAQGEDLGPVSTRA 1800
             RFGGPLY YANKDFFLQSLYSLTGLDAGCMAARRKEGGSETGVGEGGPAQGEDLGPVSTRA
Sbjct: 541    RFGGPLY YANKDFFLQSLYSLTGLDAGCMAARRKEGGSETGVGEGGPAQGEDLGPVSTRA 600

Query: 1801   ALVPAAAGFHTVVIDCAPLLFLD AAGVSTLQDLRRDYGALGISLLLACCSP PVRDILSRG 1980
             ALVPAAAGFHTVVIDCAPLLFLD AAGVSTLQDLRRDYGALGISLLLACCSP PVRDILSRG
Sbjct: 601    ALVPAAAGFHTVVIDCAPLLFLD AAGVSTLQDLRRDYGALGISLLLACCSP PVRDILSRG 660

Query: 1981   GFLGEGPGDTAEEEQFLSVHDAVQTARARHRELEATDVHL 2103
             GFLGEGPGDTAEEEQFLSVHDAVQTARARHRELEATD HL
Sbjct: 661    GFLGEGPGDTAEEEQFLSVHDAVQTARARHRELEATDAHL 701 (SEQ ID NO:4)
```

**Hmmer search results (Pfam):**

Model	Description	Score	E-value	N
PF00916	Sulfate transporter family	405.6	4.7e-118	1
CE00008	E00008 GUANYLIN	8.6	0.016	1
PF00497	Bacterial extracellular solute-binding prote	4.4	0.57	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF00497	1/1	338	356 ..	1	27 [.	4.4	0.57
CE00008	1/1	409	431 ..	1	24 [.	8.6	0.016
PF00916	1/1	195	505 ..	1	328 []	405.6	4.7e-118

```

1 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
51 NNNNNNNNNN NNNTGGTGAA ACCCGTCTC TACTAAAAAT ACAAAAAATT
101 AGCCGGGCGT GGTGGCGGGT GCCTGTAGTC CCAGCTACTC GGGAGGCTGA
151 GGCAGGAGAA TCACTTGAAC CCGGGAGACA GAGCTTGAGC TGAGCCGAGA
201 TCATGCCACT GTACTCCAGC CTGGGCAACA GAGCGAAACT CCGTCTCAAA
251 AAAAAAAAAA TTAGCCGGGC GCGGTGGCGG GCGCCTGTAG TCCCAGCTAC
301 TCAGGAGGCT GAGGCAGGAG AATGGCGTGA ACCCAGGAGG CAGAGCTTCC
351 AGTGAGCCGA GATCACACCA CTGCATTCCG GCCTGGGTGA CAGAGCAAGA
401 CTCCGCCCTCA AAAAAAAAAA AAGAAAAGGT GGGGGGCGTC TCACTATGTT
451 GACCAGGCTG GCTTTGAAC TCTGGCCTTA AGCGATCCTC CTGTCTAGGC
501 CTCCCAAAGT GTTGGAAATTA CAGGAGTGAA CCATCGTGCC TGGCTAATAA
551 TTCCTTTTAA AAAGCAGCTT ACCCTTATTT TCACGTGTGG GCCTAATTTA
601 GTTCACTTAA AAAAATCATT TATCTTCACC CCAGCCCTAT GAGGCAGGCA
651 CTGCCGGTCC TGGTCTGTGG TAGAGGGGAG GGCAGAGGAG CCGTGAGGGT
701 GACCAGGCGC TGTGGGTCCG TGCTGGGTCC AGTCAGACCA GGA CTCTGG
751 CCAGTCACGG CACCTTGACC CCGGCAGTCC TCGCCCTGGG CCGTGAGCAC
801 CACACACAGG GCTTACGCGA GCACACACGC ATATGCACGC ACCGGCAGCC
851 TTGGGCTGAG CCGGCTGTCA GCCTCTGCCC TGCTCCAGCT TGGACCAGGC
901 TGGCTCCTTG CAGGACCAGG AGGGTGTCCG GCGACTGGAC ACGGAGACCA
951 AGCCTCCCTC AGCCCGCCT GGGTTTGAAG GCTGCTGCAC TCGACCCAG
1001 ACCCCAGAGC TGAAGTTTA CCTGTGCTCA GCCCTGAGC CCCCCTCC
1051 CGCTGGTCCC TAAGCCCCC CGGCAGGGCC GCAGAGCCAC AGCTGCAGCC
1101 GCTCCTGGGA GGTCTGGAGC TCCTCAGAGG CCCACACAGC TCTAACTACT
1151 ACAAGCCCTT GATTACAGTT CAACTCCCGG ATCAGCCGAT CAGGTAACAT
1201 GGCTGGAGAA ACCCGTGA CTGCAATCTG TAGGTAAATA ATTGAAC TAC
1251 AGAGTCCAGG GCACAGACCA CTGCCTGCAG GTTGGCGCCA CCACCCAC
1301 TCTCCCGCT GCTCGCGGGA GCCAGAGGGC CCTGCGGTCC TCGGTGGTCT
1351 TGCCAGCCCC TCGTCATCCC AGGGCCCTCC GCGCTGTGA GGA CTCTCC
1401 AGGTAAGAAC CATCTGGGC CCAGATCTCA GCTGCAGCAG AGGGGGCGT
1451 GGGAGCCGAG GCCAGAAATG CCCTGGACTC GTGGTTTCTT AGGGGCACC
1501 TCAGGCTCAA GGCAGGTGGC CCTACTGTCC CCATTCCACA CACCTGGACC
1551 CCAGGGGCTT GGGGTGGGCT TCAGGGCATC CAGGGACCCA GTGTGGTGGG
1601 GTCTTCCAGG GAAGGGGACA CAACTCTTGC AATGTTGCCT GAGGGCCAGG
1651 ACCCCCGCTC TGTGCCCCAG GGGTGTGTG CCCAGCCTGC ATGTGTCAAC
1701 CTACCAGGCT GGGCTCACTG CCCCACACA CCCGCCAGGA GACTGGAGCT
1751 CGCACACCTT GGGCCAGCGT GCAAACAGCA GGCTCAGCCC AGGCTCCAGG
1801 GTGTCTGGG CACCTGGTGT CCTGGGAGCA AAGTCTTTC CTAACGTCGC
1851 TGAGAAGAAT GTTTAAAGTG AAAGTACATT GGAGTCTGCA AACAGGACAG
1901 ACCCGAGGCC TCACGTGGGA CCAGTCAGGC CTCTAAGCAC CGCCTCCCTA
1951 ACGCCACGGT GTTTTCCGAG ATCAAGGGAA AGGTGAGGTG CCCTTCCGGC
2001 TCTGCCGGCC CAGGGTGA CTGTGCAGCG GGCTGGGCCC TCTCGGTGCT
2051 GCCTCGGGAC AGTGTGTCTT GGCCGTTCCT CAGTGAGCTG GTGCAGCCTG
2101 GGAAAAAGGG CGCCTCACGT CCCAGAACTG TCTGGGCAGG GGAGACAGAC
2151 GCCAGTCAAC CTCTCCCTT CCCAGCTGGC CCTGATGGGG CCCCCGTCCA
2201 GGCATATTCT CAGAATTCTG TCCCAAGTCC AGGCGGATGG GCTAGGCTAG
2251 TGTCTGAGTG CTGCTCCCCC AGCAGACTTG GGGTCCCAGT ACCCACAAG
2301 CTGGGCAGGG ACATAGGAGG CCTCTTCTG AACTTCCGC CAGCCCCAGG
2351 ACCCACAGGG CAGGTGACAG AGGGGTGGGT GGAGGTGTCT CCAGGAGAGC
2401 AGGCGATGTT TTGGATGGGG GAGGGAGGGC TCTGTTGTGG GCATGGGGTG
2451 GACAGCAGGA CCGTTTGCCA ACCTGGGGAG CCAGGGAGGT GGACACGGAG
2501 CAGCTGGACT CAGGCTTGCC TGCACCTGTG TCCAGTGA CTGTGACATTCT
2551 GACGGTAGGC ACATGTGCGT GGTGGCAGCC CAGCCTGTTC CTGCCCCGTT
2601 GGGAGGTTGA GCTTCAGGAG GCTACAGGGT GGTTTTCAGC CAGGAACCGC
2651 AGAGCCAATA GGCCGGAGCT GAGCCTGGAC AGGGTGCCGC CACGCCGCC
2701 CTCAGCACTG CTGGCCTCAG CACACCCCAT GGCATGGGCT TGGTGTCTGA
2751 ATCCCATCTC ACCCCACGAT GGATTCTGGA TCCAGCAGGG CCCAGCGTCC
2801 ATCCATACCG GGCAGGGGGC TGGGGCCCGC GCTGCCAGGA GAAGGCCAG
2851 CACCAATCCC CGGCCCTGGG TGGGCGAGGG GTCCGCCCCA AGGGGCCCGT
2901 TGCTGCCGGG GACCTTGTCT TTTGGCCCTG GATCCGGGGG CTCCTGTGAC
2951 CATGCCCTCT TCTCGGCCGC AGGTCCGGCA CCGGACCTGA CGCAACAGGA
3001 TGGACGAGTC CCCTGAGCCT CTGCAGCAGG GCAGAGGGCC GGTGCCGGTC
3051 CGACGGCAGC GCCCAGCACC CCGGGGTCTG CGTGAGATGC TGAAGGCCAG
3101 GCTGTGGTGC AGCTGCTCGT GCAGTGTGCT GTGCGTCCGG GCGCTGGTGC

```

FIGURE 3, page 1 of 4



3151 AGGACCTGCT CCCCGCCACG CGCTGGCTGC GTCAGTACCG CCCGCGGGAG  
3201 TACCTGGCAG GCGACGTCAT GTCTGGGCTG GTCATCGGCA TCATCCTGGT  
3251 GCCGCAGGCC ATCGCCTACT CATGTCTGGC CGGGCTGCAG CCCATCTACA  
3301 GCCTCTATAC GTCCCTTCTT GCCAACCTCA TCTACTTCCT CATGGGCACC  
3351 TCACGGCATG TCTCCGTGGG CATCTTCAGC CTGCTTTGCC TCATGGTGGG  
3401 GCAGGTGGTG GACCGGGAGC TCCAGCTGGC CGGCTTTGAC CCCTCCCAGG  
3451 ACGGCCTGCA GCCCGGAGCC AACAGCAGCA CCTCAACGG CTCGGCTGCC  
3501 ATGCTGGACT GCGGGCGTGA CTGCTACGCC ATCCGTGTCT CCACCGCCCT  
3551 CACGCTGATG ACCGGGCTTT ACCAGGTGAG GAGCCCTGCT TGGGCACAGG  
3601 GAGGGGCCCC CCCCACCCCC CCTTAGGTTT TGGCCATCCA CGAGGGCAAG  
3651 GCTGGGGGCA AGCACAGGCT TGGCAGAGGA GGTGCTGGCC CAAGACAGCA  
3701 AGGCTTGGGC AGAGCTGGGG CGTGCCGGGG CATCCCAGGG CGAGGCACCG  
3751 ACGCGGAGAG GCTGTGGATG CAGGAGGGGA GGGGCACGGG GAGCCAGTCC  
3801 GGTGGGCCAT GGCCTTGGTG GGGACCAGCA GGCCAGGTGT GGCTGTGGCT  
3851 CAGTGGTGCT GGA CTGAGGC CATGTGGCCT CCCAGGCCTT CTGTCCTAGG  
3901 TGGAGTGGGG GATGGCCTCC CCACCCCCGA AGGTCTCCTG CCTTGGCCTG  
3951 TCCACCTTGG CCCCCTTGG CTCCACATCT GCATGGGGGG CAGTGGGCAC  
4001 CATGTGTAGG AAGCAGCAGG AAGGGGTTGC CTTCTGATAC CAGAGGTCTT  
4051 AATTCTGAAA TAAACGGGC TGCTGCACGT GACAAGGGTT AGACGTGTCT  
4101 ATGGCCAGCT GTGTGCACGT GTGATGCTCA CGTGGATGTC ACAGTTGTCT  
4151 GCGGGCATGA GCACGCTGG AACCAGAACT CAGGCCCGTG TGAGGAGTCT  
4201 GGT TTGGAAC ACACGGGGCC GCAACACAGA ATTGT CAGGT CCTGTGCCGT  
4251 GACCACCACC CCTCGGGCCA TGCCAGGTGC TGGTGAGGGG CAGGTGGCTC  
4301 CCGCCAGGCG CCTGCTGGCC TGACCGCACT CCGTCCACAG GTCCTCATGG  
4351 GCGTCTCCG GCTGGGCTTC GTGTCCGCT ACCTCTACA GCCACTGCTC  
4401 GATGGCTTTG CCATGGGGGC CTCCGTGACC ATCCTGACCT CGCAGCTCAA  
4451 ACACCTGCTG GCGGTGCGGA TCCCGCGCA CCAGGGGGCC GGCATGGTGG  
4501 TCCTCACATG GCTGAGCCTG CTGCGCGGCG CCGGGCAGGC CAACGTGTGC  
4551 GACGTGGTCA CCAGCACGGT GTGCTGGCG GTGCTGCTAG CCGCGAAGGA  
4601 GCTCTCAGAC CGTACCGAC ACCGCCTGAG GGTGCCGTG CCCACGGAGC  
4651 TGCTGGTCAT CGTGGTGGCC AACTCGTGT CGCACTTCGG GCAGCTCCAC  
4701 AAGCGCTTTG GCTCGAGCGT GGCTGGCGAC ATCCCCACGG GTTTCATGCC  
4751 CCCTCAGGTC CCAGAGCCCA GGCTGATGCA GCGTGTGGCT TTGGATGCCG  
4801 TGGCCCTGGC CCTCGTGGCT GCCGCCTTCT CCATCTCGCT GCGCGAGATG  
4851 TTCGCCCAGA GTCACGGCTA CTCTGTGCGT GCCAACACAG AGCTGCTGGC  
4901 TGTGGGCTGC TGCAACGTGC TACCCGCCTT CCTCCACTGC TTCGCCACCA  
4951 GCGCCGCCCT GGC CAAGAGC CTGGTGAAGA CAGCCACTGG CTGCCGGACA  
5001 CAGCTGTCCA GCGTGGTCAG CGCCACCGTG GTGCTGCTGG TGCTGCTGGC  
5051 GCTGGCACCG CTGTTCCACG ACCTACAGCG AAGCGTGTG GCCTGCGTCA  
5101 TCGTGGTCAG CTGCGGGGG GCCCTGCGCA AGGTGTGGGA CCTCCCGCGG  
5151 CTGTGGCGGA TGAGCCCGGC TGACGCGCTG GTCTGGGCAG GCACCGCGGC  
5201 CACCTGTATG CTGGTCAGCA CAGAGGCCGG GCTGCTGGCT GGCGTCATCC  
5251 TCTCGCTGCT CAGCCTGGCC GGCCGCAACC AACGCCACG CACCGCCCTG  
5301 CTGGCCCGCA TCGGGGACAC GGCTTCTAC GAGGATGCCA CAGAGTTCGA  
5351 GGGCCTCGTC CCTGAGCCCG GCGTGCGGGT GTTCCGCTTT GGGGGGCCGC  
5401 TGTA CTATGC CAACAAGGAC TTCTTCCTGC GGTACTCTA CAGCCTCACG  
5451 GGCTGGACG CAGGCTGCAT GGCTGCCAGG AGGAAGGAGG GGGGCTCAGA  
5501 GACGGGGGTC GGTGAGGGAG GCCCTGCCCA GGGCGAGGAC CTGGGCCCCG  
5551 TTAGCACCAG GGCTGCGCTG GTGCCCGCAG CGGCCGCTT CCACACAGTG  
5601 GTCATCGACT GCGCCCCGCT GCTGTTCTTA GACGCAGCTG GTGTGAGCAC  
5651 GCTGCAGGAC CTGCGCCGAG ACTACGGGGC CCTGGGCATC AGCCTGCTGC  
5701 TAGCCTGCTG CAGCCCGCCT GTGAGAGACA TTCTGAGCAG AGGAGGCTTC  
5751 CTCGGGGAGG GCCCCGGGGA CACGGCTGAG GAGGAGCAGC TGTTCCTCAG  
5801 TGTGCACGAT GCCGTGCAGA CAGCACGAGC CCGCCACAGG GAGCTGGAGG  
5851 CCACCGATGC CCATCTGTAG CAGGGCCAGG CCTGCCACG AGCCTCTGCT  
5901 CCCTCCTGGG GACCCACAGC AGACGTCTGC AAGCCACTGC TGAGACCCTT  
5951 CCAGGGAGG AGCCACCCAA GAGCTGCACT CTTGTGCCAC AGCTGCCCTG  
6001 GGGAAACCGG GGAACCCCAA CTGGGAAAGG AGGCCCTCTG ATCACACGCA  
6051 GGACCCAAAC ACTCAGAAAT CAAGAACCTC TGCCCTCCGAG ACAGGCTGGC  
6101 CCACAGTGCT GGCTGGGCCC CAATGCACCG TCCCTCAGCT CAGAAGGGAT  
6151 GGGCCTGACC TGACGCTCAG GGTGACATC TTATTTGAAC AAGGGTCCCC  
6201 CGCCATCATG CAGCCTCCAA GGTGCCAAGA GGA CTCCCTA TGCCAGGCC  
6251 TGCCCGGTGC CCACCCTGCT GGTAGGAGCC AGCGGCTCTG GCCAAGTGCA

FIGURE 3, page 2 of 4

6301 CGAGGGTCTC TGTGTTTCCA GAAGGCCCCA CACACCCAAG TGCCCCCTCAC  
6351 ACCTCGTGCC TCCCCCTCAC AGGGTGGCCA CCTGCACCAG CGTCAGGGCC  
6401 CAGGGTGTGT TGACCGATGA GACCTCAGCT CAGCCCTCAG GTGCAGTGGC  
6451 CCTACCCAGC CTGGCCAGCA GACACACACA GGGATGCTCA CGGGTGCAACC  
6501 AGGAGCCAGG TGCGGCGCAG CCAACCCTGA GCCTGCAGGG AGACCTGCAG  
6551 GAAGCCCACC GTGCCCCATG CAGGGGCTCC CTCCAGCACA CAGCCCTCAC  
6601 CCCAGCACAG CCAGCAAGGA CACGCTCTCC CCAACAGGGT GCTTCGGCGG  
6651 GAGGTGGGGG AACAGGGGT CTTCGAGCA GCCCCAGCC CTCCCCTCCC  
6701 ATCTGTGCCT CTGTAAGGGG CTCTGGGACG CCCAGACCCT GCCCGCCGCC  
6751 AACCTGGTGG TGACAAGCTC CAGCAGCCAG TGGGTCCGGA CCTGCTTGAT  
6801 GCCGCGGTGA GGGACGGCGC CCACATAGGC GAGGTTGAGC TGCTGGTCCC  
6851 AGCTGAGGAC GTACTGGTCA GCCTGGCTGT GTGGCAGCGG GGGGCTGGGG  
6901 ACAACAAAGG GCGGCTCAG TCCCGAGCCT CAGCATGGCT GGCAGCGCGG  
6951 CTGACACACA CGTTCAAGCC CAGGACTGCC CGGGCGCAGG ATCCAGGCGC  
7001 TGCCCGTGCG TTCAGTGA CTAAAAATGA CCCTTAGGGC CAGGAATGTG  
7051 GGGAGGTCCC ATCTTCATGG GGAACGGCAG CAGCAGTAAG ACGAGGGGCC  
7101 AACGCCAGCC CTGGCCCTGG CCCTGCCAGG AAGGCGGGTA CCTCAGCTCT  
7151 AGGTGGAAGG AATGGGACAG GCAGGCCAGG TCCCGCTGCA GGGCCGTCCA  
7201 CTCCAGGGG AGACTCCTGG TTTACCTCAA AGAGCAGGAT CCCGGGCATC  
7251 GGCCTGGGCT GCAGGGGGCG GCCCAGGCTC ACGCCCCGGC GCCCACTCAG  
7301 GTGGAGGACC CACCACAAA CACGGCGGGG GCGGGGCCCG GGAGAGCCAG  
7351 GGCCCCAGAG GAGGGAGCTC CGGTCTCTGA AGCTCTCACA GTGCGCAGTC  
7401 AGGGGGCGCG CGAGCTCTCC CCGTGCGGCC AGGGGGTCCC GGAGGCCGCG  
7451 GAGCGCTCAC CAGAAGCCTG TGCTCCTCCA GAAGCGCCG AGGGGCCACA  
7501 GCGCGCGGGC CGCGTCCACC TGCACCAGGT GCGGGGCCTC GGCCGGGGCC  
7551 ACCGGGGGTG CGGCCAGGAG CGAGGCCAGG AGCGCCAGCA GCGCTGCGCG  
7601 GGGGCGCAGG GGACGCATGG CCACGCGTGC TCGGGGACTG CGGGGCTTCG  
7651 GGCTGCACTG CCGGTTCGCG CTCGGGGTCG GAGTCTGGGC GCGACCCCCA  
7701 TGTGACCGCC GCCGCGGGG GGGGCCTTGG TGAGGGGGCG ATGGCCGGGT  
7751 GGGAGGGGTT GGGTGGCCTC GGGGAGCCTC GGGGAGCCG GAGCACGGCA  
7801 GGGCTTGGAG CCCCCTTCC TTGCGGGCCT CAGGGGCTGC TCTGAGGACC  
7851 GATGACTCGG AAAGCGCTCA GAAGAACGCT TCGCCCGTTG GTGCTATGTG  
7901 AGTTGAGCCA TTA CTGTCTT GTTTTCTCT GTTTTGTGT GTTTTGTGAG  
7951 CAGAGTCTTG CTTTGTGCGC CAGGCTGAGG TGCACTGGCG CGATCTCAGC  
8001 TCACTGCAAC CTCCATCTCC GGGGCTTCAG CGATTTTCTC ACCCCAGCCT  
8051 CCTGAGTAAA GCGTGCCTT TAGCAGGAAG GAGAATTACC CCAGAAGAGC  
8101 ACATTGTATA AAAATAATAA TGGAAGAAGC AGGCTTGGTG CGGTGGCTCA  
8151 AGACTGAAAA TATTGTTTAA AAGCCAAAGC AATACGAAAT AATACATATT  
8201 TTAAACAAT ACAGTATAAC AGCTATTTAC AGAGCATTTA CATTGTTTTA  
8251 GGGACTATAA GTAATCTTGA TTTAACTAC ACAGTAGGAT GTGCGTAGGT  
8301 AATGTGCAAA TACTGTGCCA TTTTATATCA AGTACTTGAG CACCTGCAAA  
8351 TTTTGGTATC TGGGAGGGTC CTGGAACCAA TACCCGAGG ATACCATGGG  
8401 ACAACTGTAG TACATGTGTA GTCCATGTAT GCATGTGTGA ATCCAAGCAA  
8451 ACATTGTATA AAAATAATAA TGGAAGAAGC AGGCTTGGTG CGGTGGCTCA  
8501 CACCTGCAAT CCCAGCACTT TGGAATTGCA GGCCAACACG GGAGGATCAC  
8551 TTGAGGCCCTG GAGTTTGAAA TCGGCCTGGG AGATGTACCA AGACCCCATC  
8601 TGTACAAAAA AAAAATTTAG CCAGATGCGA TGGTATATGC CTGTGAGGCC  
8651 CAGCTACCCA CGAAATTGAG GTGGGAGATT GCTTGAGCTT AGGAGTTCAA  
8701 GGCTGAGACG GGCCATGATC ACACCACTAC ATTCCAGCCT GGTGACAAA  
8751 ATGAGACCCC ATCTCTAAAA AAAGAAAAGA AAAAAAGAAG AGTCTACTAA  
8801 CAAAACGAAA ATACTGGACA ATAATCCTCT CTAAGTTGGG AGAAGGATAA  
8851 TTAGAGTTAC AGTGTTCCT (SEQ ID NO:3)

# FEATURES:

Start: 3000  
Exon: 3000-3575  
Intron: 3576-4340  
Exon: 4341-5867  
Stop: 5868

One non-coding Exon in the 5' UTR:  
(query = cDNA sequence; subject = genomic sequence)  
Score = 174 bits (88), Expect = 7e-46  
Identities = 91/92 (98%)  
Strand = Plus/Plus

```
Query: 1   tcgcgggagccagagggccctgcggtcctcggtggtcttgccagcccctcctcatcccag 60
          |||
Sbjct: 1313 tcgcgggagccagagggccctgcggtcctcggtggtcttgccagcccctcgtcatcccag 1372
```

```
Query: 61   ggccctccgcgcctgtgaggactccctcaggt 92
          |||
Sbjct: 1373 ggccctccgcgcctgtgaggactccctcaggt 1404
```

**CHROMOSOME MAP POSITION:**  
Chromosome 4

**ALLELIC VARIANTS:**  
C/G nucleotide polymorphism at genomic position 1363 (in non-coding exon; see  
cDNA/genomic sequence alignment above for the non-coding exon)

V/A amino acid polymorphism at protein position 699